Collaboration networks in a French cluster: do partners really interact with each other?

Bastien Bernela, Rachel Levy

Collaboration networks in a French cluster: do partners really interact with each other?

Bastien Bernela*, Rachel Levy**

*Ph.D. student in Economics, CRIEF, Poitiers

bastien.bernela@univ-poitiers.fr

**Assistant professor in Economics, LEREPS, Toulouse

rachel.levy@iut-tlse3.fr

Abstract: We discuss the common hypothesis that, in collaborative projects, all partners interact with each other in homogeneous ways. More precisely, this research aims to determine the existence and frequency of partner interactions in a collaborative project. From a survey of participants involved in innovation projects approved by a cluster, we collect information about 754 collaboration ties. We then test the impact of several determinants on the existence and frequency of their observed interactions.

Key-words: Collaboration tie, interaction, inter-organizational networks, cluster, complete graph.

JEL: L52, O32, R58

1. Introduction

According to Kline and Rosenberg's (1986) model, knowledge creation and innovation can be viewed as collective processes. Many scholars have followed this idea and focused their analysis on the interactions between heterogeneous actors in order to better understand how to foster innovation and so, indirectly, promote economic growth.

The literature has examined several questions about processes of interaction between actors. Some scholars (for example, Cassiman and Veugelers, 2002) have focused on the determinants of collaborations between heterogeneous actors in order to create knowledge, notably by focusing on partners' access to complementary knowledge or the division of research costs and risks between them. Other studies have tested the factors explaining partner choice in innovation, and particularly partners' spatial and organizational characteristics (Ferru, 2014), while another group of authors have used the CIS survey to focus on the impact of collaborative processes on innovation success (Cassiman and Veugelers, 2002; Belderbos et al., 2004; Mora-Valentin et al., 2004). More recently, in a context of increasing collaboration, many scholars have focused on how collaborative networks function, examining the idea that knowledge not only spreads via direct ties between partners, but also indirectly within innovation networks (Cowan and Jonard, 2009). Focusing particularly on the spatial dimensions of collaboration networks, recent works on the geography of innovation have mobilized network analysis as a research tool to understand better how networks function (Boschma and Frenken, 2010). In parallel with this recent interest in network studies, other (but also the same) authors have analyzed the idea that innovation creation is a localized process, and particularly takes place within clusters, i.e. "geographic concentrations of interconnected companies, specialized suppliers, service providers, firms in related industries, and associated institutions (for example, universities, standards agencies and trade associations) in particular fields that compete but also cooperate" (Porter, 1998, p.197). The literature has underlined how this notion of 'the cluster' includes both the concepts of network and of geographical proximity.

In terms of these two research areas - collaboration networks and clusters - a vast literature has developed on the analysis of innovation projects conducted within the framework of clusters. In their introduction to a special issue of Papers in Regional Science related to these subjects, Brenner et al. (2011) underline three challenges in studying knowledge networks within clusters. The first is to understand the role of clusters in the creation of innovation and economic value; the second consists in identifying the different driving forces that operate within clusters; and the third area of interest deals with the methodological challenges involved in better understanding the emergence and the structure of knowledge networks within clusters. In line with this last challenge, this study aims to test whether *complete graph* representations of scientific and innovative networks are as accurate as they are assumed to be in the empirical analyses reported in the great majority of the literature. We also develop an ordered *probit* model with selection equation to identify the determinants of interaction practices between actors involved in a collaborative research project.

In France, clusters have mainly been implemented through the government's Competitiveness Clusters (CC) policy[1]. Empirically, we focus on the case of projects conducted in a specific French CC, collecting data from the results of an online survey addressed to all the partners involved in collaborative projects 'labeled'[2] by this particular cluster. The survey, which was based on a sample of 88 collaborative projects conducted between 2006 and 2012, asked partners involved in innovation projects to assess the frequency of their interactions with every other partner, from which we collected information

[1] We use the term Competitiveness Cluster to translate the French term "pôle de compétitivité", used to designate the French government's cluster support policy (http://competitivite.gouv.fr/home-903.html). These clusters aim to reinforce the competitiveness of territories and of cluster members.
2 We use this term as reflecting the French term used in this context throughout this article: it can be seen as equating to 'approved'.

about 754 ties linking two partners. The survey provides original data, as the nature of the collaboration ties is described by the actors themselves.

The article is structured as follows. In the second section, we present the methodological background about innovation collaborative networks and their determinants, and then introduce our main hypothesis. In a third section, we present our data. We then focus on the main results about actors' observed interactions within the collaboration network, and the factors explaining them, distinguishing between their existence (fourth section) and their frequency (fifth section). We conclude by discussing methodological and policy issues about networks.

2. Empirical issues about network building

2.1. Identifying collaborative innovation processes

From the empirical articles about collaborative innovation processes published over the twenty last years, we can identify several types of data used. Interactions between firms and/or research labs within innovation networks can take different forms and can be measured by different indicators, in particular alliances (Gay and Dousset, 2005; Stuart et al., 2007; Gilsing et al., 2008), co-authorship in scientific publications (Ponds et al., 2007; Fritsch and Kauffeld-Monz, 2010; Hoekman et al., 2010), co-patenting (Hussler and Ronde, 2007; Carayol and Roux, 2008; Breschi and Lissoni, 2009; Hanaki et al., 2010), European Programs (Breschi and Cusmano, 2004; Roediger-Schluga and Barber, 2008; Autant-Bernard et al., 2007b) and research consortia (Busom and Fernandez-Ribas, 2008; Cassiman et al., 2010; Vicente et al., 2011), or even PhD students co-supervised between science and industry sponsors (Levy, 2005; Bouba-Olga et al., 2012). Some authors build their own data, mainly using case studies or surveys (Boardman and Bozeman, 2006; Arvanitis et al., 2008; Bekkers and Freitas, 2008; Cassiman et al., 2010), while others use existing datasets, including international surveys as the European community innovation survey (Cassiman and Veugelers, 2002; Belderbos et al., 2004; Mora-Valentin et al., 2004).

Regarding the diversity of data that can be used to study collaboration, we have to pay attention to the heterogeneity of the research object. First, the level of analysis can be inter-organizational or inter-individual. Levy et al. (2009) propose differentiating individual actors (e.g., publications), and structures (European programs, for example), and also between public and private actors. We must also consider carefully the type of ties: are they bilateral (as, for example, in co-publication) or unilateral (as in the case of service provision)?

The studies cited above have often specific sectoral and/or territorial delimitations, such as university-industry linkages in Austria (Schartinger et al., 2002), Switzerland (Arvanitis et al., 2008) and France (Ferru, 2014), Texas air quality research collaborations (Boardman and Bozeman, 2006), New-Zealand biomedical collaborations (He et al., 2009), or European biotechnology (Gay and Dousset, 2005; Stuart et al., 2007), nanotechnology (Autant-Bernard et al., 2007b; Cunningham and Werker, 2012), IT industry (Hanaki et al., 2010), or GNSS sector studies (Vicente et al., 2011; Balland, 2012; Balland et al., 2013), etc.

We focus in this article on the common participation in innovation projects within the framework of a French cluster. While several authors have focused on French CC policy data (Grandclement, 2011; Levy and Talbot, 2014), there is a significant difference when studying innovation partnerships between focusing on projects that are submitted and those which are actually funded and effectively realized. Some researches –such as Autant-Bernard et al. (2007b) or Grandclement (2011) – use data from project proposals rather than those that are actually accomplished, so that partner collaboration (which clearly doesn’t happen in projects that don’t actually take place) is under-represented. Since we aim to measure the existence and frequency of such collaboration, we concentrate in this article on projects that achieve funding and are actually realized.

2.2. The determinants of knowledge exchange

Before moving to network issues, we consider the determinants of partner interaction as identified in the literature. Numerous authors have tried to identify the factors which might favor collaborations between innovation actors, whether science-industry linkages or inter-firm collaborations. Many of the studies previously cited - working on patents, publications or common participation to European Framework Programs, (Autant-Bernard et al., 2007b; Hussler and Ronde, 2007; Ponds et al., 2007; Carayol and Roux, 2008; Gilsing et al., 2008; Levy et al., 2009; Fritsch and Kauffeld-Monz, 2010; Hoekman et al., 2010) - use social network analysis to test the impact of different forms of proximity on innovation activities. It is widely assumed that proximity between partners - whatever its form or definition - has a positive impact on their likelihood to interact and to innovate: “the more proximity between actors (in whatever form), the more they interact, the more they learn to innovate” (Boschma, 2005, p. 15).

Concerning different forms of proximity, despite the wide diversity of proximity grids that have been developed (Boschma, 2005; Torre and Rallet, 2005; Bouba-Olga and Grossetti, 2008, Boschma and Frenken, 2010), researchers generally agree that a basic distinction can be made between spatial and non-spatial proximity. Authors from the French school of proximity distinguish precisely between geographical and organizational proximity (Kirat and Lung, 1999; Gilly and Torre, 2000; Torre and Rallet, 2005; Bouba-Olga and Grossetti, 2008). **Geographical proximity** refers to “the spatial separation and the links in terms of distance”, and can be measured by physical distance or by localization in the same administrative territory (cf. Cunningham and Werker, 2012). In terms of non-spatial proximities, the French school defines organizational proximity as “the economic separation and links in terms of production organizations” (Gilly and Torre, 2000, p.12-13), with the idea that having the same organizational structure facilitates partner collaboration. The

nature of partnership can be measured by separating SMEs from big firms (Levy and Talbot, 2014), or by distinguishing between science-industry linkages and intra-industrial links (Cunningham and Werker, 2012). Finally, the existent literature underlines the role of **acquaintanceship** and social embeddedness (Granovetter, 1985) as drivers of interactions: in this context, Boschma (2005) uses the concept of social proximity to refer to the climate of trust between actors that can facilitate collaboration. Other studies (e.g., Gulati, 1995; Hagedoorn, 2006; Thune 2007; Ferru, 2014) have demonstrated empirically the importance for innovation projects of the reactivation of previous collaborations. We propose to test these three types of determinants on the existence and the frequency of interactions between collaborators (cf. section 3.4).

2.3. Moving from raw data to networks: the hypothesis of the *complete graph* representation

Independent of the question of the nature of data (as considered in section 2.1), we also focus on methodological issues associated with network analysis studies. Indeed, as Vonortas (2013) recalls, “in network analysis the researcher must deal with subtle issues” (p.604), the most significant of which concerns unipartite network representations and their construction from empirical data, where he notes that “the quality of the results is as good as the data they depend on” (ibid., p.604). The majority of empirical studies we find in the literature are based on a dominant - and widely accepted - hypothesis that we want to test in our survey: the *complete graph* representation (Breschi and Cusmano, 2004; Roediger-Schluga and Barber, 2008; Autant-Bernard et al., 2007b; Balland, 2012, Vonortas, 2013), which holds that all the partners involved in a collaborative innovation project will interact with each other. To build a global network representation from project data, most authors follow Breschi and Cusmano (2004) in transforming bipartite (or 2-mode) network projections - which link actors to the projects in which they are both involved - into unipartite (or 1-mode) projections linking together pairs of actors involved in the same project, as presented in Figure 1. Roediger-Schluga and Barber (2008) develop a similar method about R&D projects, and make "the assumption that the contract data produces networks that reasonably approximate to actual patterns of interaction". We could easily expand the list of references to witness the methodological dominance of what has come to be called the *complete graph*.

[INSERT FIGURE 1 HERE]

To improve this hypothesis, some authors have introduced measures to value the intensity of the ties between different network actors. For example, Roediger-Schluga and Barber (2008) define tie intensity (or weight) as the number of projects in which two partners are involved together. Other authors (e.g., Autant-Bernard et al., 2007b; Balland, 2012; Balland et al., 2013) propose eliminating occasional participations (called "alibi partners"), and considering only partners involved in at least two projects, an hypothesis that leads to what we propose to call a *multi-collaboration graph*: for example, Autant-Bernard et al. (2007b) use the joint participation of firms in a minimum of two 6^{th} Framework Program projects. However, while this methodology can be used to value ties between actors, it does not measure the frequency of their interactions within a specific project, but only how often the partnership has been renewed over time. Ties can be valued in different ways. Tie intensity can be measured using, for example, the impact factor of journals (or the number of citations of an article in scientific publications, as in He et al., 2009), or the funding associated with a contract (Busom and Fernandez-Ribas, 2008). In this article, we propose to study the intensity of interactions within innovation collaborative projects by using a quantitative measure of their frequency (cf. section 3.3).

Other works (Breschi and Cusmano, 2004; Grandclement, 2011) suggest that the coordinator of the project (who Breschi and Cusmano call the "prime-contractor") is connected to every participant via their dominant position, but without observing ties between the other participants, a hypothesis that leads to a *star graph* representation, and tends to overestimate the strength of ties involving the coordinator relative to those involving other partners.

Finally, some other authors build networks between cities or regions by aggregating the local structures within those areas (see Scherngell and Barber, 2009 on the 5th European Framework Program; Maggioni et al., 2007 on co-patenting; Ponds et al., 2007 and Hoekman et al., 2010 on scientific publications; Bouba-Olga et al., 2012 on co-supervision of PhD students). These works focus on the geographical dimensions of collaboration, and use a gravity model to identify the determinants of partners' spatial distributions. Here, tie valuation refers to the number of partnerships between two territories over a given period, whereas we want to measure inter-organizational ties within the framework of a collaborative project: as we study collaborative ties rather than territory pairs, such models do not align with our research.

2.4. Applying social network analysis

The studies on innovation collaborative processes noted above have usually been conducted in combination with some recent advance in sociological theory, and particularly social network analysis: indeed, Autant-Bernard et al. (2007a) refers to "the networked nature of knowledge creation and the geography of innovation". Their objective is to identify, from the network structure and the actors' positions, the best ways to foster innovation (Hussler and Ronde, 2007; Balland, 2012; Balland et al., 2013).

Many scholars apply the methodology of social network analysis in order to build what they call innovation networks (or knowledge diffusion networks), from which they propose to characterize network structures and compute indicators of actors' (i.e. nodes') positions within networks. More precisely, network structures can be characterized by their size (numbers of nodes and ties), their density (numbers of actual ties divided by the total possible number of ties), their connectivity (number and size of major components and number of isolated nodes) and geodesic distances (the shortest possible path between two network nodes). Finally, there can be indicators about the degree of clustering (presence of 'grapes') within the network, usually measured by the number of order triples that are transitive (Borgatti et al., 2002). For example, Breschi and Cusmano (2004) characterize European program networks by their density, the number of components involved, the size of the two largest components (the core of the network), the degree of clustering, the average and maximum distance between any two nodes, and the average degree centrality of the nodes within the largest component. Roediger-Schluga and Barber (2008) also take into account the entry and exit of actors into and from the network between different periods.

Three indicators of actor centrality are usually employed to measure actors' positions within global networks (Borgatti et al., 2002; Borgatti and Foster, 2003): *degree centrality* (i.e. the number of ties linking a node to other network nodes); *closeness centrality* (a measure of the distance between one node and other network nodes); and *betweenness centrality* (a measure of an actor's intermediary position between other network nodes), usually used to measure the level of control exercised by that actor over network activity (Levy and Talbot, 2014). Autant-Bernard (2007b) use actor's network positions to measure the social distance between them.

All these indicators (including network structures and node positions) are calculated under the *complete graph* hypothesis. Implicit in that representation is the assumption that all actors in a network are connected to all other actors, so that knowledge is automatically diffused throughout the network via their common participation in innovative projects. But, to our knowledge, little is known about real interactions within collaborative projects, so this hypothesis lacks empirical evidence. Of course, it is almost impossible to really measure knowledge diffusion – but we propose to approximate it by measuring the frequency of interactions between network actors using survey data. Thus, the hypothesis we test in this article is that: *interactions between partners involved in a collaborative project are heterogeneous (in terms of existence and frequency).* We argue that, in reality, interactions between such partners can be better represented via an *empirical graph* - since, in practice, some ties exist and some do not, and some ties are stronger than others. Figure 2 opposes the two types of representations: the theoretical (the *complete graph*) on the left, and the *empirical* on the right. The hypothesis we seek to validate is the transition from the first representation to the second.

[INSERT FIGURE 2 HERE]

Given that some potential links between actors in innovative projects do not actually exist, and that some ties are weaker than others, we can suppose that knowledge is not homogeneously diffused within the project, and thus that the *complete graph* hypothesis is not totally valid. Literature often considers ties as supporting knowledge diffusion (Cassiman et al., 2010; Balland, 2012), and that the fact of being involved in the same project implies that actors share knowledge (automatic assumptions that Gomes-Casseres et al. (2005) discuss). Consequently, we must pay attention to conclusions driven by social network analyses which apply the *complete graph* hypothesis to partnership data. Finally, we provide an econometrical analysis of the determinants of existence and frequency of interactions between partners. The aim is to go beyond debating the *complete graph* hypothesis, and better understand why interactions within collaborative projects are not homogeneous.

3. Data and method

3.1. Case study on a French competitiveness cluster

In 2005, the French government implemented a national cluster policy to create competitiveness clusters (CCs), which it defined as “joint theme-based initiatives for a given geographical area, i.e. in a given territory, that bring together companies, research centers and educational institutions in order to develop synergies and cooperative efforts targeted at one (or more) given market(s) (…) clusters using synergies and innovative joint projects to give their members a chance to be national and international leaders in their fields" (www.competitivite.gouv.fr). In concrete terms, 71 CCs have been established within French territories (some of them globally oriented, some nationally oriented), each specializing in a general sector, such as electronics, biotechnology, wood industry, etc. Cluster members are usually located within the same NUTS2 region, but occasionally spread over two or three contiguous regions. These clusters are all organized as associations, with memberships that include several firms and research laboratories or Higher Education and Research Establishments (HEREs) located in their geographical areas and more or less concerned with their specialized sector or technology.

As well as their management and territorial marketing activities, CCs are also required to encourage the development of innovation projects, especially between cluster members, although they also often involve partners from beyond the clusters' home zones. Firms and HEREs from each cluster propose innovation collaborative projects that are launched in a two-step procedure. First, they are labeled by the CC, depending on the project's innovative characteristics and on its links with the cluster's strategy. In some cases - depending on the project's subject - they may be co-labeled by several CCs, following the second phase of the government's cluster policy (from 2010), which emphasized inter-clustering and cooperation between members of different CCs.

Once labeled, each project must find funding, which can come from various different sources: two national funding schemes – i) the first administered by the National Research Agency, and ii) the second from a governmental fund dedicated to CC projects – or from iii) European funding, generally through European Framework Programs and European Regional

Development Fund; or iv) local funding, mostly from local authorities (Regional Councils, public investment banks, etc.).

Our study uses data about projects labeled by a national CC[3] which have been run since the CC policy was implemented in 2005. By the end of 2012, this cluster had acquired 76 members and had labeled 284 projects: comparing these figures with those of others French CCs, this cluster is about average, and so represents a relevant setting in which to analyze collaboration within clusters (cf. EuroLIO, 2010 for a comparison of a sample of 20 similar CCs in terms of numbers of establishments involved and of their employees).

Among these 284 labeled projects, we focus only on projects which actually gained funding, and which have two or more partners. More precisely, we exclude the 174 projects that failed to attract funding, as some of them may have never been conducted (although some scholars do not consider this methodological precaution, and also include proposals and unrealized projects (Autant-Bernard et al., 2007b), and so cannot be sure of the actual existence of some of the projects they study). We also exclude 22 projects that only had single participants (usually start-up creations), because they do not fit with our aim to research collaboration processes.

This research is therefore based on a sample of 88 collaborative projects conducted between 2006 and 2012 and involving 262 different establishments (firms or HEREs), each of which participated (on average) in 1.8 projects (participations per establishment ranged from 1 - 22). In other words, the data we collected represents a total sample of 475 project participations.

3.2. Variables on projects and partners

Table 1 describes the data used in this empirical research, detailing some descriptive statistics about participation in projects, some of which relate to the projects, and others to the partners

[3] This CC wants to remain anonymous, so this article does not refer to anything that could allow it to be identified.

involved in them[4]. We use the following information to characterize projects:

- **Project size** (*project_size*), which is defined by the number of partners involved. The projects studied had between 2 and 19 partners, with a mean of 5.4 partners. In what follows, we use this mean to distinguish two sizes of projects to simplify our analysis: small projects (with a maximum of 5 partners) vs. large projects (with 6 or more partners).
- **Funding** (*funding*)**.** As noted above, we can distinguish four forms of funding: two national forms: i.e. from the National Research Agency (*research_agency*) and from the governmental fund dedicated to CC policy (*cc_policy*); as well as European funding (*europe*); and local funding (*local*).
- **Co-labeling** (*colabeling*)**.** We record information about the co-labeling of a project, i.e. when it is approved by at least two CCs.
- **Year of labeling** (*period_label*). The 88 collaborative projects we study were labeled between 2006 and 2012. As with project size, we simplify our analysis by distinguishing two distinct periods: the first phase of the CC policy (*period1*) refers to projects labeled between 2005 and 2009, and the second (*period2*) to those labeled between 2010 and 2012.

We also use some data about the partners participating in the sample projects:

- **Coordinator** (*coordinator*)**.** Each project is led by a coordinator, the establishment that is the driving force behind the project and ensures the smooth running of the collaboration.
- **Local.** We consider that a partner is local when located in the CC's own area.
- **Member.** Establishments can participate in CCs without actually being cluster

[4] Table 1 includes a representativeness test of the survey's respondent population that we comment in the following section.

members, so we record information about the cluster membership of each project participant.

- **Structure.** We distinguish three types of structures: HEREs, SMEs (<250 employees), and groups (of larger establishments).

[INSERT TABLE 1 HERE]

3.3. Measuring partners' interactions within a collaborative project: a survey method

The object of this article is to provide a critical assessment of the dominant theoretical hypothesis about the nature of collaboration ties within networks. In order to get qualitative and declarative data about participants' actual interactions during real-life projects, we decided to conduct an online survey (Eisenhardt, 1989; Yin, 2003). The survey was addressed to all partners involved in the 88 targeted collaborative projects labeled by the CC (i.e. 475 participations), and sent to their referents as noted in the CC's mailing list. We asked them to answer from the framework of a specific project and describe their interactions with all other project partners. The survey was sent by email in early June 2013, and two follow-up emails were sent to non-respondents after a two-week interval. The CC director sent further follow-up emails to cluster members who had still not replied, and we finally closed the survey in mid-July.

Before studying the response rate of the survey and the representativeness of the respondent sample, we give some information about the content of the survey. To try to achieve a high response rate, the survey was very short: in fact, it contained only two questions. Bearing our hypothesis in mind, the first question concerned the frequency of the respondent's interactions all other project partners. Following quantitative survey methodology (Eisenhardt, 1989; Yin, 2003), we proposed a scale of four different frequencies of interaction, as well as of the absence of interaction – i.e.:

- 0: No interaction at all
- 1: Very few interactions, i.e. less than once a year
- 2: Few interactions, i.e. more than once a year but less than once trimester
- 3: Regular interactions, i.e. more than once a trimester, but less than once a month
- 4: Very regular interaction, i.e. more than once a month

The choice of this scale was motivated by the fact that all projects lasted at least one year, so each possible answer would make sense over that timescale. To ensure the scale was reliable, we discussed and validated it with the CC director. We then added a second question about the partners' previous acquaintance before the project, asking if they knew each other and if they had worked together before, with the object of collecting relevant empirical material to build a variable about acquaintanceship (cf. section 3.4).

Of the 475 project participants, we actually only sent 371 surveys, as 104 referent email addresses were missing, and finally collected 186 responses, i.e. a 50% response rate of surveyed partners (and 39% of those initially targeted), which is satisfactory. We ran chi-2 tests on each descriptive variable to check bias relative to all responses to find out whether the respondents' profiles differed from those of the overall targeted population (cf. Table 1). We found that local partners, CC members and project coordinators were comparatively overrepresented in the sample of respondents[5]. We can legitimately assume that these actors were more receptive to our research because of their stronger links to the CC: moreover, the CC director sent follow-up emails to cluster members, which probably increased their response rate. The number of partners involved in projects also appears to have impacted their likelihood to answer the survey, as the response rate for small projects was higher. Apart from this overrepresentation - which did not modify the interpretation of our results - there

[5] These three variables were correlated (at 0.1%): more than 90% of members were local, and coordinators were members in 45 of the 88 projects surveyed. We therefore chose to focus on the '*coordinator'* variable and excluded *local* and *member* from our analysis.

was no bias in responses according to respondents' structures, or type of structure, or funding, or period of labeling, or co-labeling. We also note that Table 1 shows that HEREs are overrepresented in the population cluster studied in comparison with others (EuroLIO, 2010) - and as a consequence - the proportion of projects funded by the National Research Agency is also greater.

3.4. The particular unit of analysis: the tie between two partners

To analyze the interactions between partners in innovation projects, we focused at the level of the tie between two partners participating in the same project, and tested the impact of the characteristics of projects and of partners on the existence and frequency of their interactions. This choice of tie analysis requires prior disambiguation - since, for the same tie, we can get two answers describing the level of interactions. For instance, if actors A and B are involved in a same project, A can describe its perceptions as to its level of interactions with B, and B of its interactions with A. Thus, there are two possible responses about the same tie, and these two responses may differ. In terms of the existence of interactions, of the overall total of 197 ties for which we got two answers (from A and B), in 3.5% of the cases one party reported zero interaction, while the other reported interaction at some level. Concerning the frequency of interactions, we got different answers from the two partners for about 62.4% of the ties, but in most cases (78.1%) those differences represented offsets of only one degree of frequency. Heterogeneous interpretations from the two partners as to interaction frequency in same tie are thus low. When the responses from the two partners differed, we used the lower frequency as our measure for that variable. We justify this choice in the following way: as our research question concerns the issue of collaboration intensity, we consider that, when two partners have different perceptions of the intensity of their collaboration, taking the greater perceived intensity into account would risk overestimating the real intensity.

The 186 responses to the survey represent 754 different ties. Table 2 describes the composition of this analysis sample using the explanatory variables defined in section 3.2. Using the tie as the unit of analysis also allows for the introduction of the following variables to better define the relationship between the two partners:

Geographical proximity (*geo_proxi*)**.** We build a binary variable to define geographical proximity: two partners are considered geographically close if they are located in the same or the neighboring (NUTS3) area. In terms of French geography, this criterion is relevant at the scale of the NUTS3, the NUTS2 region being significantlty larger. We use the place where the effective project work took place, rather than its administrative HQ, as the partner's geographical location.

Nature of the partnership (*partnership*). Following Levy et al. (2009), we distinguish between collaborations where the partners are public and private actors. From data about their type of structure provided by the CC (HEREs, SMEs, groups), we differentiate three types of partnerships. First science-science ties (*sc_sc)* involve two HEREs; second, industry-industry ties (*ind_ind*) involve two SME(s) and/or group(s); and third, science-industry ties (*sc_ind*) involve a HERE and an SME or group.

Acquaintanceship (*acquaintance*)**.** Previous partnership is a crucial point when studying collaboration, but this information is often the most difficult to assess. One of the main interests of our research is to survey partners directly about their previous acquaintance and to test the impact of this variable on interactions within collaborative projects. We asked each partner to report if they knew or had previously worked with the other before the focal project, and used their answers to measure acquaintanceship as a binary variable.

[INSERT TABLE 2 HERE]

4. Results about the existence of interactions between partners

4.1. Descriptive statistics about the existence of interactions and their determinants

One of the main objects of this article is to empirically test the dominant representation of collaboration networks, called the *complete graph* - and its underlying general hypothesis that every partner in a collaborative project interacts with every other partner - by simply looking at the existence of interactions between all the partners of innovative projects in our case cluster.

Table 3 shows that some ties (48 out of 754) are characterized by the absence of any interactions between partners, but that the great majority (93.6%) of ties involve interactions, a result that tends - at least partly - to support the *complete graph* hypothesis (which assumes all ties are interactive). However, as the following section shows, these non-existing ties can change some interpretations about innovation network structures. Table 3 gives some descriptive statistics about interactions and the results of a *binary probit*[6] test showing the impact of explanatory variables on the existence (or not) of interactions between two partners in collaborative project ties.

[INSERT TABLE 3 HERE]

Table 3 identifies three main results. First, we observe the negative and significant impact of European funding on the existence of interactions: more than one fifth of such projects (21.4%) are characterized by the absence of interactions between partners. One possible explanation may involve the matter of project coordination. European projects involve an average of 11.8 partners per project, against an average of 5.4 for all the projects studied and, given this higher number of partners in such consortia, European projects are often run as sets of sub-projects linked together via a coordinator (Breschi and Cusmano, 2004), so that partners may well not all interact directly with each other. To our knowledge, this result has not been previously noted in the literature, and calls for some restraint vis-à-vis the use of the

[6] We test a simple *binary probit*: the variable explained is the existence of an interaction, which takes the value of 1 if interaction exists, 0 otherwise.

complete graph hypothesis in analyzing large European projects.

Second, Table 3 highlights the significant role of the coordinator: respondents reported no interactions between the two partners in only 1.5% of ties involving coordinators. The *coordinator* variable is highly significant in explaining the existence of interactions between partners, and this marginal effect indicates that being project coordinator increases the probability of interacting with other participants by 5.3%. This result demonstrates the central position of coordinators, and confirms the legitimacy of the *star graph* representation. While the likelihood that interactions take place between coordinators and other partners is very high, there are also many ties between the other partners. As a consequence, we can see the *empirical graph* as an intermediary between the *complete* and the *star graph* representations: interactions exist in the ties between most partners involved in a collaborative project, and particularly in those involving the coordinator.

Finally, an interesting result concerns the positive and significant impact of previous acquaintance between actors on the probability that they interact: partners who have experience of collaborating together (whatever the form of their previous collaboration) are more likely to interact during their focal project. Of the 754 ties recorded, 539 (71.5%) were characterized by previous acquaintance between partners, which underlines the importance of social proximity as a modality of linkage and as a determinant of the likelihood of future collaborations between the partners (Gulati, 1995; Grossetti and Bès, 2003; Boschma, 2005; Hagedoorn, 2006; Thune, 2007; Ferru, 2014).

Several of the variables we included in our analysis seemed to have no significant impact on the likelihood of observing interactions between partners: this was particularly the case for project size[7] and co-labeling. Nor did geographical proximity of partners impact the probability that they would interact, echoing results by Bouba-Olga et al. (2012) who found a

[7] We also tested the size as a continuous variable, and it is not significant either.

mitigated impact of geographical proximity on collaboration (depending on regional characteristics), and Cunningham and Werker (2012) who also found that geographical proximity (as measured by partners' presence in the same administrative region) had a non-significant impact on their likelihood to interact. A possible explanation is the existence of temporary geographical proximity between the two partners (Torre, 2008; Torre, 2011), allowing partners who weren't actually co-located to meet once or twice during collaborative projects. Another variable that appears to have no impact on the probability of partner interaction is the nature of the partnership[8], i.e. the type of partners involved. Being from different worlds (from science or industry) seems to have no significant impact on how smoothly the project runs, a result that also confirms Cunningham and Werker's (2012) findings.

4.2. Comparison of theoretical and empirical networks

In the previous section, we identified the number of ties where there was no interaction, which is quite low (6.4%). We now aim to assess the impact of these few missing ties on the network properties, comparing structural characteristics and positions of actors between theoretical and empirical graphs.

As explained in section 3.3, we could not collect information about the (non)existence of interactions between partners across the whole network, since our final response rate was 39%. Not having a 100% response rate means we cannot build a full empirical graph representation for the whole network, and we therefore introduce two hypotheses in order to build two estimates (a lower and a higher) of partner interaction rates from the 754 ties for which we did get answers. In our *low hypothesis*, we suppose that ties between non-respondent partners are characterized by an absence of interaction: only ties for which we get a positive answer confirming the existence of interactions are represented. In contrast, in our

[8] We also tested this variable by distinguishing SME from group, and it is again not significant.

high hypothesis, we suppose that interactions did exist between non-respondent partners: so only ties where the actors involved confirmed the non-existence of interactions in their answers are regarded as non-interactive. We propose comparing the *complete graph* with the *high hypothesis* empirical graph - which takes the most optimistic stance towards the existence of partner interactions - in order to estimate the impact of the removal of a few non-existing ties. Table 4 presents some indicators that compare the structure of these two graphs (which are illustrated in Appendix A).

[INSERT TABLE 4 HERE]

In Table 4, and in these two graphs, we observe no significant differences in terms of density and average geodesic distance, which confirms the result presented in the previous section, and tends to validate the *complete graph* hypothesis for representing the network of actors inside a cluster. Nevertheless, we note a difference in the graphs' diameters (maximum geodesic distance), which underlines how the presence of some weak ties increases the connectivity of the network (Granovetter, 1983 or Friedkin, 1982).

As well as looking at the network structures, we can also compare the two graphs by looking at the position of actors inside the network. Table 5 shows the top ten network nodes in terms of normalized centrality following the three classical modes of calculating centrality as previously defined (degree, closeness, and betweenness) between the two.

[INSERT TABLE 5 HERE]

We use a Kendall rank correlation test to check whether the differences observed are significant: the result reveals they are not, whichever mode of calculation is used, confirming the partial validity of the *complete graph* hypothesis. Nevertheless, although the differences are not significant, we do observe some differences between the two rankings, particularly concerning betweenness centrality, the position that allows actors to control knowledge diffusion across the innovation network (cf for example Levy et Talbot, 2014). While the two most central actors are the same for the theoretical and empirical graphs, actor A49 (a technology transfer center located in the same administrative region as the CC) is in the fifth position for the *complete graph*, but is the third most central actor in the *high hypothesis graph*. Thus, including non-existing ties when applying the *complete graph* hypothesis decreases the intermediary role of this actor, which is responsible for transferring technology between partners.

These results about the different rankings in nodes' positions, as well of the difference in the network diameters, confirm the idea that even if the *complete graph* hypothesis is partially acceptable we must be cautious, as the deletion of some ties can change the characterization of the network, and thus the diffusion of knowledge within it. These non-existing links could be the weak ties which allow a network to be more cohesive.

In the same way that the inclusion of non-existing ties may change a network's characterization, the removal of existing links can also modify it. If we compare the *star graph* and the *multi-collaboration graph* (represented in Appendix B), we can note that these two forms of representation strongly underestimate the total number of existing network ties compared to the *low hypothesis* graph: indeed, the *star graph* shows 362 ties, and the *multi-collaboration graph* only 74, as against the 675 represented in the *low hypothesis graph*. This result reinforces the necessity of comparing *complete graphs* with other types of representation in order to discuss the empirical choice of graph representations before drawing hasty conclusions.

5. Results about the frequency of interactions between partners

5.1. The model

We have shown that interactions within collaborative innovation projects are not perfectly homogeneous in terms of existence. Table 6 reports partners' answers about their interactions and also demonstrates that interactions are not homogeneous in terms of frequency either. On the scale detailed in section 3.3, we observe that 85 partner couples interacted less than once a year, whereas 200 interacted more than once a month.

[INSERT TABLE 6 HERE]

We want to determine the factors explaining this heterogeneity of interactions. Our object is to test econometrically the impact of different determinants on both the existence and frequency of interactions in a collaborative project. We therefore ran an *ordered probit* with a sample selection to identify which factors (primarily, measures of proximity) could explain first the existence of interactions between two partners, and second, their frequency. The variable to be explained is discrete and ordered, and data observability is restricted by a binary selection mechanism (De Luca and Perotti, 2010). The introduction of a selection equation allows the potential bias of the existence of interactions to be taken into account before studying their frequency.

An ordered response model with sample selection can be represented by the following bivariate threshold crossing model:

$Y_j^* = \beta_j X_j + \mu_j \; with \; j = 1{,}2$ **Equation (1)**

$Y_1 = I(Y_1^* \geq 0)$ **Equation (2)**

$Y_2 = \sum_{h=0}^{H} h \, I\left(\alpha_h < Y_j^* \leq \alpha_{h+1}\right) \; if \; Y_1 = 1$ **Equation (3)**

where Y_1^* and Y_2^* represent continuous latent variables for the selection process and the outcome of interest respectively, the β_j are k_j vectors of unknown parameters, the X_j are k_j vectors of exogenous variables, and μ_j represents random errors (Eq. (1)). The latent variable Y_1^* is related to the binary indicator Y_1 through the observational rule (Eq. (2)), and I(A) denotes the indicator function of the event A. The latent variable Y_2^* is related to the outcome Y_2 through the observational rule (Eq. (3)), where $\alpha = (\alpha_1, \ldots, \alpha_H)$, with $\alpha_h < \alpha_{h+1}$, $\alpha_0 = -\infty$ and $\alpha_{H+1} = +\infty$ is a vector of H with strictly increasing thresholds which partitions Y_2^* into H+1 exhaustive and mutually exclusive intervals. As in a classical sample selection model, the observability of Y_2 is confined to the sub-sample of observations for which $Y_1 = 1$ (the selected sample). Selectivity effects operate via the correlation between the latent regression errors μ_1 and μ_2.

In the selection equation Y_1 (Eq. (2)), which concerns the existence of interactions between the two partners, the explained variable takes the value of 1 if there are interactions and 0 if the partners do not interact. In the outcome equation Y_2 (Eq. (3)), which concerns the frequency of such interactions, the explained variable takes the value 1, 2, 3 or 4 according to the frequency scale used in the survey. We use the same explanatory variables in the two equations.

5.2. The determinants of frequency of interactions

Before commenting on the results presented in Table 7, we must note that the inclusion of a selection equation, legitimized by the nature of our dependent variable, does not introduce bias (*rho* is not significant). In other words, the results obtained in Eq. (1) correspond perfectly with those obtained from the *binary probit* test in section 4.1; while an *ordered probit* test without the selection equation would have given the same results as Eq. (2). For each explanatory variable, the results differ depending on whether it has a significant impact on both the existence and frequency of interactions, or on only one of the two, or on neither.

[INSERT TABLE 7 HERE]

The frequency of interactions between partners is independent of the project size: when a consortium is composed of more partners, they do not appear to interact more or less frequently. With regard to the type of funding, we observe significant and negative impacts on the frequency of interactions for projects funded by European programs and by the French national research agency. For European projects, we have already commented on their tendency to adopt sub-project structures, which may explain why some partner couples in the same project do not necessarily interact at all. This variable also has a significant but lesser impact on the frequency of interactions. Concerning the national research agency, the fact there are less frequent interactions than in CC policy projects can be explained by the fact that the CC's interventions mean the projects are more structured. In fact, CCs give their projects a great deal of support and attention, as they act as a showcase for the cluster's identity and success: this support tends to reinforce the levels of interaction between partners.

Co-labeling has a negative impact on the frequency of interaction between partners. Co-labeling – where project partners are members of both the CC we studied and of other CCs - have a negative impact on the frequency of partner interactions. In 2009, the French government encouraged inter-clustering - i.e. collaborations between partners from different CCs - and this policy orientation appears to have had a real influence, as the proportion of co-labeled projects subsequently increased from 27% to 35%. While such an increase could have been obtained by artificially linking some actors to build inter-cluster projects, it mainly represents new collaborations in which players first have to get to know each other: that may explain the lower levels of interaction in such projects than in non-co-labeled projects, in which partners are more likely to benefit from previous acquaintance.

The role of coordinator also seems to be important in projects: the frequency of interactions is significantly higher for ties involving coordinators, confirming previous results and supporting the need to combine the *star graph* and *complete graph* representations. Geographical proximity has no effect on the existence of interactions, nor on their frequency. If temporary proximity is not always the explanation, actors can also use ICT to interact at a distance (Cairncross, 1997; Morgan, 2004; Charlot and Duranton, 2006; Aguilera and Lethiais, 2011). Nor does nature of the partnership (science and/or industry) effect the existence or frequency of interactions, confirming Cunningham and Werker's (2012) results. Finally, previous acquaintance between partners has the highest and positive coefficient in the model, highlighting the importance of social relationships and mutual confidence in supporting coordination.

6. Conclusion

The main objective of this study is to gather empirical declarative data to better understand the nature of interactions in collaborative innovation projects. Our research was based on a survey addressed to partners involved in a French competitiveness cluster, from which we gained information on the existence and frequency of interactions between partners in 754 collaborative ties. This case study is not intended to make judgments on the French national cluster policy, but rather to learn from original and current material from this source.

Regarding our study's results, we can consider that *complete graph* representations improperly assume on average 7% of ties to be active. This amount may seem at the same time both negligible and decisive: negligible because it represents a small proportion of the total links, encouraging the validation of the *complete graph* representation; but decisive because network properties can be easily disrupted by the deletion of only a few strategic ties. So we recommend being cautious about the use of *complete graphs*: even if our study demonstrates that the *empirical graph* – as obtained from the actors' own declarations – is not very different from the *complete graph*, it confirms that interactions between partner couples are far from being homogeneous.

In terms of the determinants of interaction, we observe the following three variables have stable and significant impacts on both the existence and frequency of interactions. First, interactions are less likely to exist and are more infrequent in European projects than in other projects. This result is especially important, given the huge literature focusing on European framework programs: applying the *complete graph* hypothesis to these projects would definitely be unwise, and one can question the encouragement of the construction of large consortia in which partner interactions seem to be more difficult. Second, coordinators generally appear to have important structuring roles in projects: on average, partners interact more with them than they do with the other actors. While ties involving coordinators are not the only ones that exist, they are usually significantly stronger. This result legitimates the underlying idea of *star graph*, but - as we demonstrate - this representation risks underestimating the number of existing ties: superimposing the *star graph* on the *complete graph* would give a more accurate representation of the weight of network ties. Third, previous acquaintance between partners is the most significant determinant of the frequency of their interactions during collaborative projects. Having previous collaborative experience facilitates the operation of the current project, supporting arguments about the importance of sociological dimensions. In terms of policy implications, the main objective of clusters is to expand their networks and gain new members, but this finding suggests they should also focus on consolidating existing ties based on social relationships, and stresses the benefits of meetings organized within the CC framework which encourage actors to meet and exchange in informal ways.

While this article introduces methodological insights into the network analysis of collaborative innovation projects, several limitations need to be noted. We face the traditional disadvantage of survey research - the incompleteness of answers - so that our data does not cover the whole network of the studied CC. Comparison with other CCs, and controlling by sectoral specialization and location, could consolidate the findings of this research.

Moreover, a surprising result is that partners' geographical proximity does not play a significant role on the existence and frequency of their interactions. A useful further step in considering project coordination would be to distinguish between face-to-face interactions and those that occur over a distance. Despite our findings, one can assume that geographical proximity would have a positive and higher impact on face-to-face interactions than on those that take place at a distance (Cairncross, 1997; Morgan, 2004; Charlot and Duranton, 2006; Aguilera and Lethiais, 2011).

Finally, we chose to use the frequency of interactions to define their intensity. This measure is more quantitative than qualitative, and further research should study the relationship between the quantity and the quality of interactions on knowledge diffusion. This issue is all the more important, given that the literature predominantly links collaborative ties with knowledge diffusion. This association can be extended to collaboration success, prompting consideration as to whether more frequent interactions make initial project ambitions more likely to be realized. In terms of policy implications, this highlights the need to combine quantitative and qualitative methods when evaluating innovation policy.

References

Aguilera A, Lethiais V (2011) Transmission des connaissances dans les relations de coopération inter-entreprises: TIC versus face-à-face. *Revue d'Economie Régionale et Urbaine* 2: 269-293

Arvanitis S, Kubli U, Woerter M (2008) University-industry knowledge and technology transfer in Switzerland: what university scientists think about co-operation with private enterprises. *Research Policy* 37: 1865-1883

Autant-Bernard C, Mairesse J, Massard N (2007a) Spatial knowledge diffusion through collaborative networks. *Papers in Regional Science* 86: 341-350

Autant-Bernard C, Billand P, Frachisse D, Massard N (2007b) Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies. *Papers in Regional Science* 86: 495-520

Balland PA (2012) Proximity and the evolution of collaboration networks: evidence from research and development projects within the Global Navigation Satellite System (GNSS) Industry. *Regional Studies* 46: 741-756

Balland PA, Suire R, Vicente J (2013) Structural and geographical patterns of knowledge networks in emerging technological standards: evidence from the European GNSS industry. *Economics of Innovation and New Technology* 22: 47-72

Bekkers R, Freitas I (2008) Analysing knowledge transfer channels between universities and industry: to what degree do sectors also matter? *Research Policy* 37: 1837-1853

Belderbos R, Carree M, Lokshin B (2004) R&D cooperation and firm performance. *Research Policy* 33: 1477 – 1492

Boardman C, Bozeman B (2006) Implementing a ‘bottom-up’ multi-sector research collaboration: the case of the Texas air quality study. *Economics of Innovation and New Technology* 15: 51-69

Borgatti SP, Everett MG, Freeman LC (2002) *Ucinet for Windows: software for social network analysis*. Harvard, Analytic Technologies

Borgatti SP, Foster PC (2003) The network paradigm in organizational research: a review and typology, *Journal of Management* 29: 991-1013

Boschma R (2005) Proximity and innovation: a critical assessment. *Regional Studies* 39: 61-74

Boschma R, Frenken K (2010) The spatial evolution of innovation networks: a proximity perspective. In: Boschma R, Martin R (eds) *Handbook of Evolutionary Economic Geography*. Cheltenham, Edward Elgar

Bouba-Olga O, Grossetti M (2008) Socio-économie de proximité. *Revue d'Economie Régionale et Urbaine* 3 : 311-318

Bouba-Olga O, Ferru M, Pepin D (2012) Exploring spatial features of science-industry partnerships, a study on French data. *Papers in Regional Science* 91: 355-375

Brenner T, Cantner U Fornahl D, Fromhold-Eisebith M, Werker C (2011) Regional innovation systems, clusters, and knowledge networking. *Papers in Regional Science* 90: 243-249

Breschi S, Cusmano L (2004) Unveiling the texture of a European research area: emergence of oligarchic networks under EU framework programmes. *International Journal of Technology Management* 27: 747–772

Breschi S, Lissoni F (2009) Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography* 9: 439-468

Busom I, Fernandez-Ribas A (2008) The impact of firm participation in R&D programmes on R&D partnerships. *Research Policy* 37: 240-257

Cairncross F (1997) *The death of distance: how the communications revolution will change our lives*. Harvard Business School Publishing, Cambridge

Carayol N, Roux P (2008) The strategic formation of inter-individual collaboration networks, evidence from co-invention patterns. *Annales d'Economie et de Statistique* 87: 275-301

Cassiman B, Di Guardo MC, Valentini G (2010) Organizing links with science: cooperate or contract? A project-level analysis. *Research Policy* 39: 882-892

Cassiman B, Veugelers R (2002) R&D Cooperation and spillovers: some empirical evidence from Belgium. *The American Economic Review* 92: 1169-1184

Charlot S, Duranton G (2006) Cities and workplace communication, some French evidence. *Urban Studies* 43: 1365-1394

Cowan R, Jonard N (2009) Knowledge portfolios and the organization of innovation networks. *Academy of Management Review* 34: 320-342

Cunningham W, Werker C (2012) Proximity and collaboration in European nanotechnology. *Papers in Regional Science* 91: 723-742

De Luca G, Perotti V (2010) Estimation of ordered response models with sample selection. *CEIS Working Paper* 168

Eisenhardt K (1989) Building theory from case study research. *Academic Management Review* 14: 532–550

EuroLIO (2010) L'innovation dans les pôles de compétitivité. (http://www.eurolio.eu/Productions/Fiche-Pole)

Ferru M (2014) Partners' connection process and spatial effects: new insights from a comparative inter-organizational partnerships analysis. *European Planning Studies*, Forthcoming

Friedkin, N. E. (1982) Information flow through strong and weak ties in intraorganizational social networks. *Social Networks* 3: 273-285

Fritsch M, Kauffeld-Monz M (2010) The impact of network structure on knowledge transfer: an application of social network analysis in the context of regional innovation networks. *Annals of Regional Science* 44: 21-38

Gay B, Dousset B (2005) Innovation and network structural dynamics: study of the alliance network of a major sector of the biotechnology industry. *Research Policy* 34: 1457-1475

Gilly JP, Torre A (2000) *Dynamiques de Proximité*. L'Harmattan, Paris

Gilsing V, Nooteboom B, Vanhaverbeke W, Duysters G, Van den Oord A (2008) Network embeddedness and the exploration of novel technologies: technological distance, betweenness centrality and density. *Research Policy* 37: 1717-1731

Gomes-Casseres B, Hagedoorn J, Jaffe A (2005) Do alliances promote knowledge flows? *Journal of Financial Economics* 80: 5-33

Grandclement A (2011) Vers une géographie des pôles de compétitivité : réseaux, construction politique et organisation territoriale. In : Resmini L, Torre A (eds) *Competitivita territoriale: determinanti e politiche*, Franco Angeli, Milan

Granovetter M (1983) The strength of weak ties: a network theory revisited. *Sociological Theory* 1: 201-233

Granovetter M (1985) Economic action and social structure: the problem of embeddedness. *American Journal of Sociology* 91: 533-544

Grossetti M, Bès MP (2003) Dynamique des réseaux et des cercles : encastrements et découplages. *Revue d'Economie Industrielle* 103: 43-58

Gulati R (1995) Social structure and alliance formation patterns: a longitudinal analysis. *Administrative Science Quaterly* 40: 619–652

Hagedoorn J (2006) Understanding the cross-level embeddedness of inter-firm partnership formation. *Academy of Management Review* 31: 670–680

Hanaki N, Nakajima R, Ogura Y (2010) The dynamics of R&D network in the IT industry. *Research Policy* 39: 386-399.

He ZL, Geng XS, Campbell-Hunt C (2009) Research collaboration and research output: a longitudinal study of 65 biomedical scientists in a New Zealand university. *Research Policy* 38: 306-317

Hoekman J, Frenken K, Tijssen RJW (2010) Research collaboration at a distance: changing spatial patterns of scientific collaboration within Europe. *Research Policy* 39: 662-673

Hussler C, Ronde P (2007) Explaining the geography of co-patenting in the scientific community: a social network analysis. *Annales d'Economie et de Statistique* 87: 351-372

Kirat T, Lung Y (1999) Innovation and proximity, territories as loci of collective learning processes. *European Urban and Regional Studies* 6: 27-38

Kline S, Rosenberg N (1986) An overview of innovation. In: Landau R, Rosenberg N (eds) The *Positive sum strategy: harnessing technology for economic growth*. Washington, National Academy Press

Levy R (2005) Les doctorants CIFRE : médiateurs entre laboratoires de recherche universitaires et entreprises. *Revue d'Economie Industrielle* 111 : 79-96

Levy R, Roux P, Wolff S (2009) Study of science-industry collaborative patterns in a large European university. *Journal of Technology Transfer* 34: 1-23

Levy R, Talbot D (2014) Control by proximity: evidence from the "Aerospace Valley" Competitiveness Cluster. *Regional Studies*, Forthcoming

Maggioni M, Nosvelli M, Uberti T (2007) Space versus networks in the geography of innovation: a European analysis. *Papers in Regional Science* 86: 471-493

Mora-Valentin EM, Montoro-Sanchez A, Guerras-Martin LA (2004) Determining factors in the success of R&D cooperative agreements between firms and research organizations. *Research Policy* 33: 17–40

Morgan K (2004) The exaggerated death of geography: learning, proximity and territorial innovation systems. *Journal of Economic Geography* 4: 3-21

Ponds R, Van Oort F, Frenken K (2007) The geographical and institutional proximity of research collaboration. *Papers in Regional Science* 86: 423-443

Porter M (1998) Clusters and the new economics of competition. *Harvard Business Review* 76: 77-90

Roediger-Schluga T, Barber MJ (2008). R&D collaboration networks in the European Framework Programmes: data processing, network construction and selected results. *International Journal of Foresight and Innovation Policy* 4: 321-347

Schartinger D, Rammer C, Fischer MM, Frohlich J (2002) Knowledge interactions between universities and industry in Austria: sectoral patterns and determinants. *Research Policy* 31: 303-328

Scherngell T, Barber MJ (2009). Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU framework programme. *Papers in Regional Science* 88: 531-546

Stuart TE, Ozdemir SZ, Ding W (2007) Vertical alliance networks: the case of university-biotechnology-pharmaceutical alliance chains. *Research Policy* 36: 477-498

Thune T (2007) University-industry collaboration: the network embeddedness approach. *Science and Public Policy* 34: 158–168

Torre A (2008) On the role played by temporary geographical proximity in knowledge transmission. *Regional Studies* 42: 869-889

Torre A (2011) The role of proximity during long-distance collaborative projects: temporary geographical proximity helps. *International Journal Foresight and Innovation Policy,* 7: 213-230

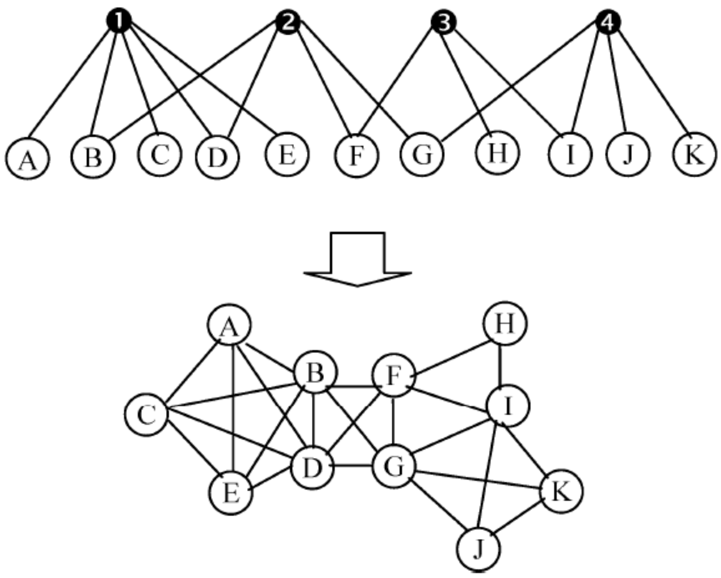
Torre A, Rallet A (2005) Proximity and localization. *Regional Studies* 39: 47-60

Vicente J, Balland PA, Brossard O (2011) Getting into networks and clusters: evidence from the Midi-Pyrenean Global Navigation Satellite Systems (GNSS) collaboration network. *Regional Studies* 45: 1059-1078

Vonortas NS (2013) Social networks in R&D program evaluation. *Journal of Technology Transfer* 38: 577-606

Yin RK (2003) *Case Study Research, Design and Methods*, Sage Publications

Figure 1: From bipartite to unipartite network ***(from Breschi and Cusmano, 2004, p.757)***



Top: Bipartite graph of organisations (A to K) and projects (1 to 4), with lines linking each organisation to the project in which it participated.
Bottom: The one-mode projection of the same network onto just organisations.

Figure 2: Representation of theoretical and empirical graphs

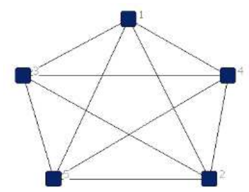
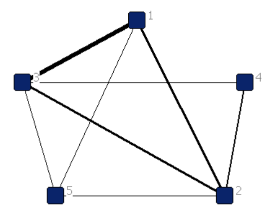
	
Complete graph (theoretical)	*Observed graph* (empirical)

Table 1: Descriptive statistics on population and respondents

		Targeted population		Respondent population		Chi2 test
		n	%	n	%	
Partner variables	*local*	166	34.9	91	48.9	***
	member	161	33.9	95	51.1	***
	coordinator	88	18.5	51	27.4	***
	structure					n.s.
	HERE	251	52.9	112	60.2	
	group	77	16.2	27	14.5	
	SME	147	30.9	51	25.3	
Project variables	*project_size*					**
	≤5	169	35.6	84	45.2	
	>5	306	64.4	102	54.8	
	funding					n.s.
	cc_policy	224	47.2	97	52.2	
	research_agency	135	28.4	50	26.9	
	europe	46	9.7	9	4.8	
	local	70	14.7	30	16.1	
	period_label					n.s.
	period1	252	53.1	100	53.8	
	period2	223	46.9	86	46.2	
	colabeling	173	36.4	72	38.7	n.s.
	Total	**475**	**100.0**	**186**	**100.0**	

**P<0.05; **P<0.01; ***P<0.001; n.s.: non-significant*

Table 2: Descriptive statistics on ties

		n	**%**
Project variables	*project_size*		
	≤5	172	22.8
	>5	582	77.2
	funding		
	cc_policy	250	33.2
	research_agency	350	46.4
	europe	98	13.0
	local	56	7.4
	period_label		
	period1	366	48.5
	period2	388	51.5
	colabeling	351	46.5
Ties variables	*coordinator*	270	35.8
	geo_proxi	217	28.8
	partnership		
	sc_sc	207	27.5
	ind_ind	198	26.3
	sc_ind	349	46.2
	acquaintance	539	71.5
	Total	**754**	**100.0**

Table 3: Descriptive statistics and *binary probit* results on the existence of interactions

		Descriptive statistics		Binary probit (n=754; pseudo r^2=0.20)
		Nb.Obs. (ties)	% with interaction	marginal effects (dF/dX)
Project variables	*project_size*			
	≤5	172	95.3	ref.
	>5	582	93.1	-0.1
	funding			
	cc_policy	250	95.6	ref.
	research_agency	350	97.1	0.4
	europe	98	78.6	-14.9 ***
	local	56	89.3	-9.5 *
	period_label			
	period1	366	94.8	ref.
	period2	388	92.5	3.3 *
	colabeling	351	93.2	-0.9
Ties variables	*coordinator*	270	98.5	5.3 ***
	geo_proxi	217	93.6	-1.8
	partnership			
	sc_sc	207	96.6	ref.
	ind_ind	198	88.9	-3.2
	sc_ind	349	94.6	0.2
	acquaintance	539	96.3	4.1 **
	Total	**754**	**93.6**	

*P<0.05; **P<0.01; ***P<0.001

Table 4: Structural characteristics of theoretical and empirical graphs

	***Complete graph* (theoretical)**	***High hypothesis graph* (empirical)**
Number of nodes	262	262
Number of ties	1335	1282
Density	1.95%	1.87%
Average geodesic distance	3.022	3.107
Maximum geodesic distance	6	7
Transitivity: % of ordered triples in which i-->j and j-->k that are transitive:	48.56%	47.71%

Table 5: Centrality indicators on theoretical and empirical graphs

Betweenness centrality			Degree centrality			Closeness centrality		
ID	*Complete graph*	*High hypothesis graph*	ID	*Complete graph*	*High hypothesis graph*	ID	*Complete graph*	*High hypothesis graph*
A267	28.55 (1)	28.57 (1)	A267	28.74 (1)	28.74 (1)	A267	51.58 (1)	49.43 (1)
A268	17.05 (2)	17.56 (2)	A49	20.31 (2)	20.31 (2)	A49	49.62 (2)	48.42 (2)
A264	15.90 (3)	15.95 (4)	A264	18.39 (3)	17.62 (3)	A31	48.24 (3)	45.95 (3)
A31	15.81 (4)	13.13 (5)	A31	16.86 (4)	16.48 (4)	A264	45.79 (4)	44.69 (5)
A49	15.53 (5)	16.97 (3)	A268	16.48 (5)	15.33 (5)	A226	45.31 (5)	45 (4)
A265	13.36 (6)	8.67 (8)	A265	16.48 (6)	14.94 (6)	A266	44.69 (6)	44.31 (6)
A266	10.08 (7)	12.04 (6)	A266	15.71 (7)	14.56 (7)	A268	44.24 (7)	43.94 (7)
A226	9.31 (8)	10.46 (7)	A89	12.64 (8)	10.73 (10)	A233	42.78 (8)	41.89 (8)
A80	8.10 (9)	8.10 (10)	A235	11.49 (9)	11.11 (9)	A240	41.89 (9)	41.03 (9)
A89	7.11 (10)	8.44 (9)	A226	11.49 (10)	11.11 (8)	A69	41.82 (10)	39.48 (16)
Kendall's rank correlation 0.786***			Kendall's rank correlation 0.825***			Kendall's rank correlation 0.919**		

**P<0.05; **P<0.01; ***P<0.001*

Table 6: Distribution of interaction frequency by ties

Scale of frequency	**n**	**%**
0: No interaction at all	48	6.4
1: Very few interactions	85	11.3
2: Few interactions	158	20.9
3: Regular interactions	263	34.9
4: Very regular interactions	200	26.5
Total	754	100.0

Table 7: Estimation of ordered probit with sample selection

		Eq(1): selection equation on existence n=754	Eq(2): output equation on frequency n=706
Project variables	*project_size*		
	≤5	ref.	ref.
	>5	-0.05	0.20
	funding		
	cc_policy	ref.	ref.
	research_agency	0.08	-0.29*
	europe	-1.09***	-0.42*
	local	-0.80*	0.07
	period_label		
	period1	ref.	ref.
	period2	0.51*	0.17
	colabeling	-0.12	-0.33**
Tie variables	*coordinator*	0.94***	0.52***
	geo_proxi	-0.24	0.07
	partnership		
	sc_sc	ref.	ref.
	ind_ind	0.04	-0.07
	sc_ind	-0.38	-0.06
	acquaintance	0.50**	0.56***
	constant	1.35***	

*Wald chi2=100.53***; Log Likelihood=-1018.35 ; rho=0.25 n.s.*

**P<0.05; **P<0.01; ***P<0.001; n.s.: non-significant*

Appendix A: Complete graph and high hypothesis empirical graph representations

Figure A.1: Complete graph representation

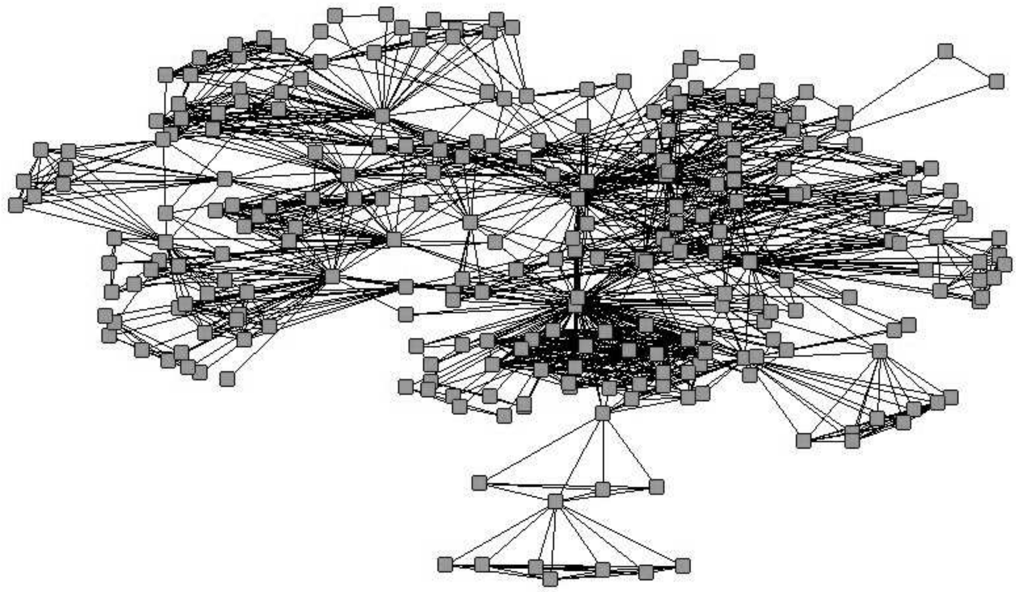
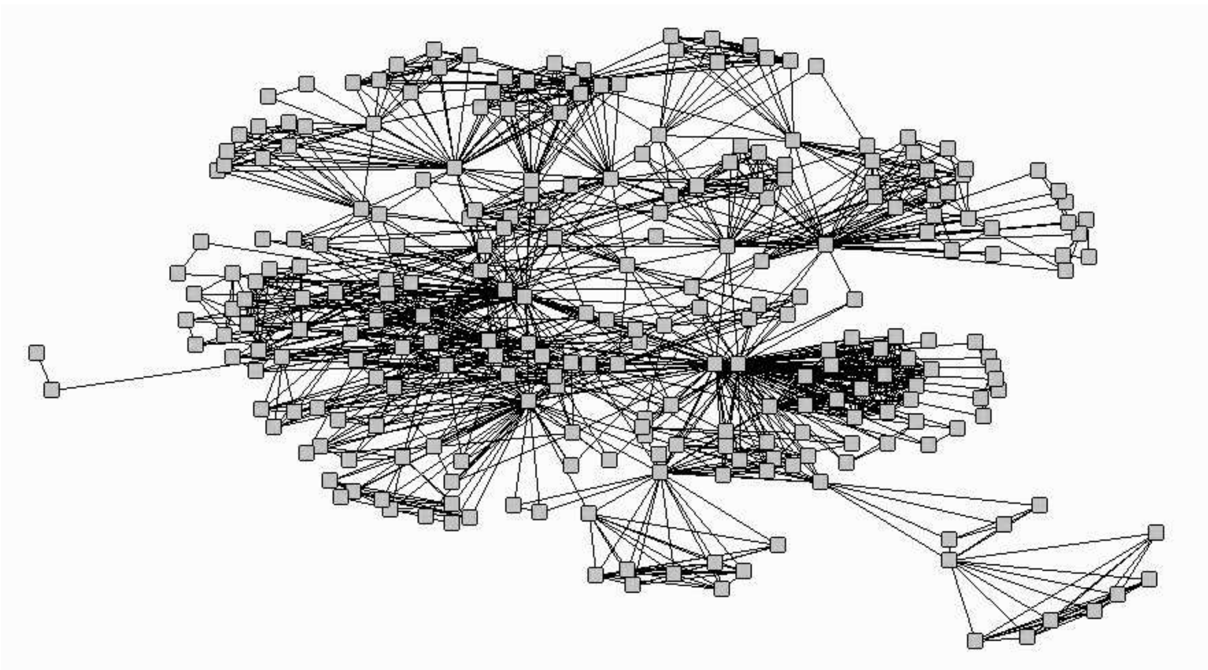


Figure A.2: High hypothesis empirical graph representation



Appendix B: Low hypothesis empirical graph, star graph and multi-collaboration graph representations

Figure B.1: Low hypothesis empirical graph representation

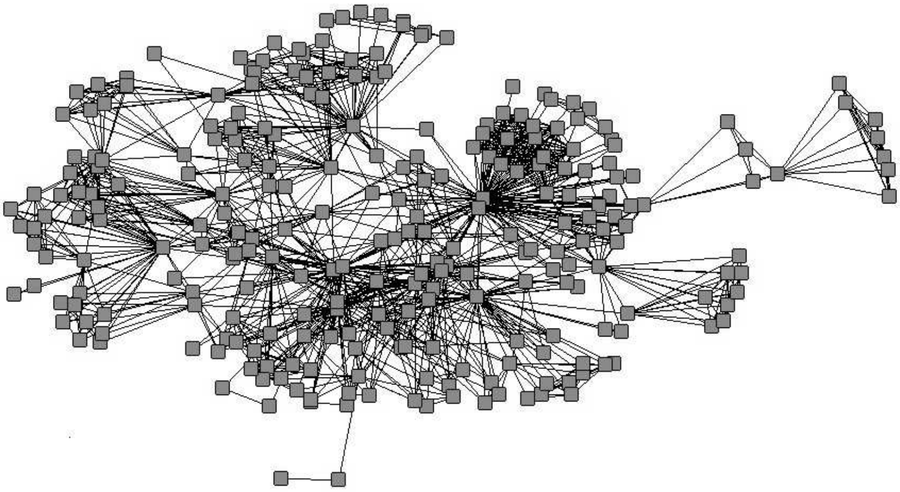


Figure B.2: Star graph representation

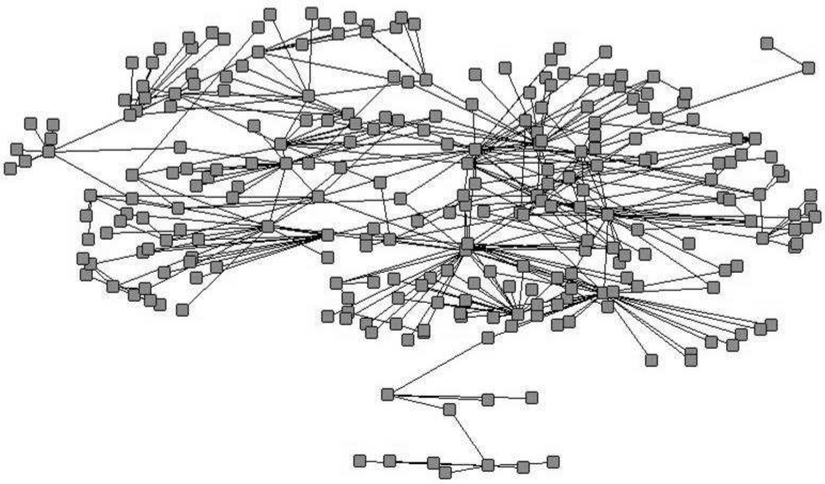


Figure B.3: Multi-collaboration graph representation